JPPIPA 11(2) (2025)

Jurnal Penelitian Pendidikan IPA

*Journal of Research in Science Education*

http://jppipa.unram.ac.id/index.php/jppipa/index

# Customer Segmentation Based on Recency, Frequency, Monetary Analysis Using K-Means Algorithms in Apple Ecosystem

Edwin Setiawan[1*], Bayu Surarso[2], Dinar Mutiara Kusumo Nugraheni[3]

[1] Magister Information System, Universitas Diponegoro, Semarang, Indonesia.
[2] Graduate School and Department of Mathematics, Univeristas Diponegoro, Semarang, Indonesia.
[3] Computer Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia.

Received: December 13, 2024
Revised: January 27, 2025
Accepted: February 25, 2025
Published: February 28, 2025

Corresponding Author:
Edwin Setiawan
mr.edwinsetiawan@gmail.com

DOI: 10.29303/jppipa.v11i2.10011



**Abstract:** One of the companies in Semarang engaged in gadget sales services has an Apple Ecosystem information system for selling products from an exclusive brand, Apple. Inside there are sales transactions and also service devices iPad, Macbook Air, Macbook Pro, AirPods, Mac, and Apple Accsessories. This research uses purchase transaction data from Apple Ecosystem customers for the period 2023. The use of RFM (Recency, Frequency, Monetary) analysis helps in determining the attributes used for customer segmentation. To determine the optimal number of clusters from the RFM dataset, the Elbow method is applied. The dataset generated from RFM is grouped using the K-Means algorithm, the quality of the algorithm will be compared in cluster formation using the Silhouette Coefficient method. All procedures will be loaded into the Customer Segmentation App (RFM Clustering) web application. Customer segmentation from RFM datasets that have been clustered produces 3 optimal clusters, namely Cluster 2 is High Spenders with 326 customers, Cluster 0 is VIP Customers, Cluster 1 is Frequent Buyers. Cluster validation of k-means using the silhouette coefficient produces a value of 0.3524.

**Keywords:** Customer Segmentation; Elbow Method; K-Means clustering; RFM Models; Silhouette Coefficient

## Introduction

In an increasingly competitive era of globalization, the intensification of competition in the business world requires effective strategies to gain an advantage (Awalina and Rahayu, 2023). Most companies have begun to utilize transaction data to be processed using a certain method, resulting in important information that can predict or build a business strategy that can provide progress to the company. The many needs of the community that must be met make companies more active in making variations in the sale of their products so that they are increasingly popular with the public. The emergence of trust in the brand makes consumers have a perception of the product, especially if the product is considered good at providing quality that satisfies consumer perceptions (Firmansyah and Widyana, 2021).

One of the companies in Semarang engaged in gadget sales services has an Apple Ecosystem information system for selling products from an exclusive brand, Apple. Inside there are sales transactions and also service devices iPad, Macbook Air, Macbook Pro, AirPods, Mac, and Apple Accsessories. An effective database management system is needed to successfully implement the company's business strategy for its customers, which stems from knowledge management and data mining techniques (Firmansyah and Widyana, 2021). Data mining techniques have made a big difference in realizing the concept of customer management in a business (Christy, et al 2021). It includes searching in databases such as knowledge

**How to Cite:**
Setiawan, E., Surarso, B., & Nugraheni, D. M. K. (2025). Customer Segmentation Based on Recency, Frequency, Monetary Analysis Using K-Means Algorithms in Apple Ecosystem. *Jurnal Penelitian Pendidikan IPA*, *11*(2), 634–641. https://doi.org/10.29303/jppipa.v11i2.10011

rules, patterns, regularities, and other patterns hidden in the data (Ikotun, et al, 2023). One of the solutions used by companies in facing challenges in this modern era is by utilizing observations and analysis of customer behavior by maintaining existing customers, because the cost of acquiring new customers will be much greater than maintaining existing customers (Hayati, et al, 2020).

An effective data management system is needed to achieve customer behavior patterns that can be seen from transaction data in the Apple Ecosystem which is very much, especially in the current industrial era 4.0. Customer segmentation is a strategy applied to group customers into different categories based on differences in their characteristics, behavior, or needs (Maskanah, et al, 2020). This approach is very important because it allows companies to use supporting data that is useful in identifying the level of customer loyalty, as well as developing effective and efficient strategies in achieving business goals (Nisa and Heikal, 2022). In the process, companies face challenges in segmenting customers due to the large amount of transaction data, so managing it manually becomes impossible due to human limitations in processing data (Kristanti, et al., 2024).

Therefore, the application of data mining becomes very useful for managing data on a large scale. Data mining is a series of processes used to find important information from large and complex sets of data that are difficult to process manually (Nikmatun, 2019). K-Means Clustering with the RFM (Recency, Frequency, Monetary) model is an effective approach in customer segmentation in business data analysis (Wijaya, et al, 2021). This research uses data mining methods with the K-Means Clustering algorithm and uses RFM data that has been processed. Apart from using RFM data, this research gets data from the Apple Ecosystem. To get the most optimal number of segmentations in K-Means, this research will use segmentation evaluation methods, namely Elbow Method and Silhouette Coefficient. The results of this study can help the company in making business decisions, especially by managing promotional costs tailored to the character of each customer segmentation so that they are right on target and can help in efforts to achieve company sales targets.

## Method

*Overall research method*

This research uses experimental quantitative research methods. The data used in this study comes from transaction data for 1 year (January-December 2023) in the Apple Ecosystem information system. This year the company experienced an increase in sales compared to the previous year, and was the highest peak moment of all YOY (Year on Year) sales from the last 5 years. In order for the data modeling process to be more optimal, the data to be processed passes the preprocessing stage and will produce an RFM (Recency, Frequency, Monetary) model. Furthermore, the dataset will be used for the modeling process using the K-means algorithm. distance calculation will be calculated using the Euclidean Distance formula, while for the process of evaluating the number of clusters it is recommended to use the Elbow Method and Silhouette Coefficient.

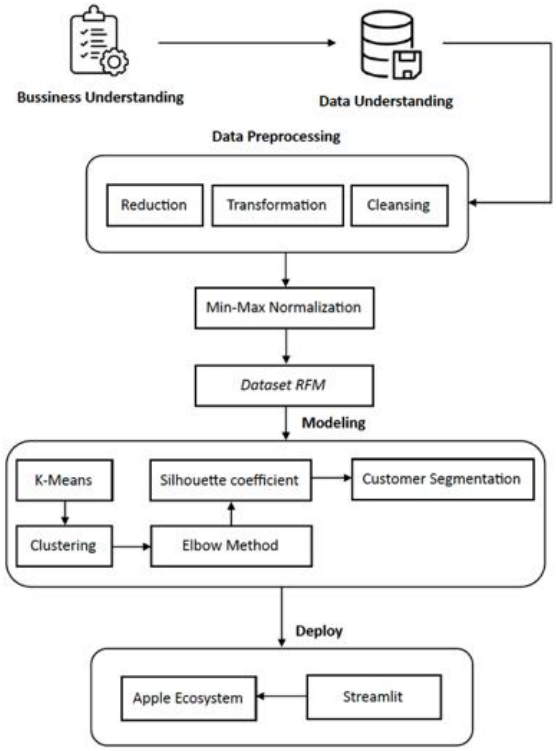


**Figure 1.** Reseacrh Procedures

*Data Collecting*

Data collection for customer segmentation analysis is obtained from product transaction data in the Apple Ecosystem which is taken in the form of purchase date, customer name, customer ID, product ID, number of product purchases, price per item, and total purchase price. In this section, researchers collect and select data based on transaction data from the period of 2023 with RFM (recency, frequency, monetary). Reference data extracted from the database in .csv format so that data pre-processing can be done.

*Preprocessing Data*

Before the clustering process, it is necessary to pre-process the data through data cleaning, outlier reduction, and transform data to transform the value range of each variable and make Min-Max normalization to change the value range of each variable to 0 to 1. The next stage researchers will carry out the clustering process with the K-means method. After that, we will determine the value of k with the Elbow method

intuitively and will carry out an evaluation method using the silhouette coefficient to determine the accuracy of data in groups or clusters. Analysis of cluster results and data will come out as output, namely the results of Apple Ecosystem customer segmentation. When the data is in accordance with the RFM format, it will then be processed on the customer segmentation app web application (RFM Clustering). Normalization is done automatically on the web application by taking the average of the initial range and the final range of each data. This data is ready to be used for the next process. The next process is the modeling process using the K-means algorithm and a series of methods to determine the number of centroids.

*RFM Model*

RFM is an analytical method to identify customer attitudes and represent customer attitudes based on 3 attributes, namely Recency, Frequency, Monetary. According to Tsiptsis and Chorianopoulos (2009), RFM analysis consists of Recency, Frequency, Monetary which has the following meanings (Bagul, dkk, 2021):

1. Recency, is a variable to measure customer value based on the time span (date, month, year) of the customer's last transaction to date. The smaller the time span, the greater the recency value.
2. Frequency, is a variable to measure customer value based on the number of transactions made by customers in one period. The greater the number of transactions carried out, the greater the f value.
3. Monetary, is a variable to measure customer value based on the amount of money issued by customers in one period. The greater the amount of money spent

The RFM concept can be used to measure customer behavior according to their previous transaction history. According to Cheng and Chen (2009), the greater the R and F values, the more likely the customer will make a return transaction with the company. In addition, the greater the value of M, the tendency of customers to respond to the company's products and services (Shirole, et al, 2021).

*Customer Segmentation*

Segmentation serves as a company tactic to gain a deeper understanding of the intended consumers, thus enabling the ongoing business to produce optimal results. Companies can meet specific customer needs and build strong relationships with customers through communication that focuses on personalization (Sarkar, et al, 2024). According to (Heri Sudarsono, 2020) in his book entitled Marketing Management states that customer segmentation can also increase customer loyalty and help companies maintain control of their market. Companies can meet specific customer needs and build strong relationships with customers through communication that focuses on personalization. According to research (Wibowo, 2021), there is a grouping of data that has the same data character and each grouped data will be analyzed in order to understand its character. In this study, customer segmentation is divided into 10 classes according to their characteristics based on RFM values as shown in Table 1.

**Table 1.** Customer Segmentation

| Customer Classs | Characteristics | Strategy Suggestions |
|---|---|---|
| VIP Customer | The most valuable and loyal customers with high transactions | Provide personalized gifts, loyalty programs, and premium customer services |
| Frequent Buyers | Customers who frequently transact, though with relatively low purchase value | Offer loyalty rewards and special promotions to increase purchase frequency |
| High Spenders | Customers with high transaction value but infrequent purchases | Upsell premium products or offer exclusive experiences |
| Potential Loyalists | Customers showing loyalty potential but not yet fully committed | Increase engagement by offering personalized loyalty promotions |
| New Customers | Customers who are still exploring and have not yet been consistent | Provide welcome offers and guide them in discovering relevant products |
| Occasional Buyers | Customers who buy sporadically or rarely | Offer targeted discounts and recommendations to encourage more frequent purchases |
| At-Risk Customers | Customers who were once active but have now reduced their activity | Re-engage with personalized offers and recovery campaigns |
| Lost Customers | Customers who were previously active but no longer make purchases | Offer bigger discounts to attract them back. |
| Bargain Hunters | Customers who are very price-sensitive and always look for discounts | Highlight special promotions and limited-time offers |
| Window Shoppers | Customers who frequently browse but rarely make a purchase | Provide targeted incentives such as discounts or personalized product recommendations |

*Elbow Method*

Elbow method is a method that aims to help find the optimal number of clusters in the dataset. To determine the optimal K value, the K value will be checked one by one and the SSE (Sum Square Error) value will be recorded. SSE (Sum of Square Error) is a formula used to measure the difference between the data obtained with the forecast model that has been done previously (Ashari, dkk, 2023). Graphical approach and old approach to find the optimal number of K in K-means algorithm. The elbow method is usually presented in the form of a graph to find out more clearly the elbow formed (Kong, et al., 2021).

The purpose of the elbow method is to choose a small k value and still have a low withinss value. The value of k in the elbow combination with k-means is a graph of the relationship of clusters to error reduction. The number of clusters k resulting from testing with k-means is evaluated with the SSE technique. SSE (Sum of Square Error) is a formula used to measure the difference between data that has been done before (Rocha, et al, 2021).

SSE is defined as follows:

$$SSE = \sum_{k=1}^{K} \sum_{x_i \in s_k} \left\| X_i - C_k \right\|_2^2 \quad (1)$$

Description:
K = number of clusters
Sk = k-cluster, xi is an element of the kth cluster
Ck = centroid of cluster Sk
||· || = Euclidean distance between two data patterns.

We should plot a line chart graph between SSE and the cluster value corresponding to K. K starts with 2 and will increase by 1 at each step. If the line chart graph shows a drastic decrease in SSE, for example, like an arm, then the "elbow" on the arm is a value that indicates the corresponding number of clusters k in K-means clustering (Alam, et al., 2024).

*Silhouette Coefficient*

Silhouette Coefficient method is one of the methods used to test the quality of clusters from the clustering process. Silhouette Index will evaluate the placement of each object in each cluster by comparing the average distance of objects in one cluster and the distance between objects in different clusters (Hidayanti, dkk, 2021).

The following is the equation used in the global silhouette (Khan, dkk, 2024):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i) - b(i)\}} \quad (1)$$

The explanation is as follows, $S(i)$ is the silhouette value. $b(i)$ is the average distance from object i to all objects in the same cluster. $a(i)$ is the average distance from object I to objects in different clusters. Based on the standardized silhouette, where a higher value is better than a lower value, a value close to zero is considered not good (Alam, dkk, 2024).

*K-Means Algorithm*

K-Means is a non-hierarchical data grouping method that divides data into two or more groups. This method divides data into several groups so that data with the same characteristics are included in the same group, and data with different characteristics are grouped into other groups (Vysala, dkk, 2020). The objective function used for K-Means is determined based on the distance and value of the data objects in the group (Putra & Wadisman, 2018). The steps in performing clustering with k-means are as follows (Triyansyah & Fitrianah, 2018):

1. Specifies the number of clusters.
2. Allocate data randomly into existing clusters according to the closest distance.
3. Calculate the average of each cluster from the data in each cluster.
4. Re-allocate all data to the cluster according to the closest distance.
5. Repeat process number 3, until no changes or changes occur.

## Result and Discussion

*Bussiness Understanding*

The data used in this research is transaction data derived from the Apple Ecosystem information system. The flow of data collection in the study is limited only to transactions for the purchase of goods for 1 year in the period January-December 2023. Because in this one year the company experienced quite fluctuating sales figures so that the data was used to find the right and efficient solution. Furthermore, the data will be processed into RFM data (Recency, Frequency, Monetary) which is expected to help adjustments in inputting data into web applications.

*Data Understanding*

This research uses secondary data that is already available in the data warehouse, in this study we will use Apple Ecosystem transaction data. The data retrieval technique is to query directly to the Apple Ecosystem information system along with the conditions needed according to the scope of the research and export the data into a .csv file. For transaction data as a whole, there are 4,100 databases. From all available data, data is

obtained by applying RFM (Recency, Frequency, Monetary). Data retrieved in the form of :
- Customer ID
- Purchase date data
- Purchase amount data
- Grand total data

*Data Preparation*

a. Data Cleansing

The data cleaning process is carried out in two stages, the first is cleaned through query conditions when data collection takes place, and the second data cleaning is through the preparation according to the RFM model.

b. Data Aggregate

From the data that has been prepared, then the aggregation will be carried out for each customer ID. The data will have Customer ID, the number of days from January-December 2023 with the last date of shopping (Recency), the number of transactions (Frequency), and the grand total amount (Monetary).

**Table 2.** RFM Data

| No | Customer ID | Transaction Date | Monetary Value | ID Transaksi |
|---|---|---|---|---|
| 1 | d9f0b7e0-e310-4094-8872-5ce0703b5520 | 17/04/23 | 763126 | 6b6bc7b0-0dd1-4512-8973-b7cb720fea11 |
| 2 | 1207fc17-3cc0-40df-93a5-a0922f942bbc | 11/08/23 | 1477497 | 61527026-f1ba-4512-a395-3c5b0ea0ac27 |
| 3 | a0431ba5-3b1c-4e16-8bf0-fd047b91a549 | 31/03/23 | 1951285 | f0e5c784-7f2e-4e10-8720-a849b909a813 |
| 4 | 2d312c78-49f8-3fa6-b49a-7ba419e71ac8 | 21/06/23 | 2038943 | 62de3f23-1c25-4a5d-9199-4f0b58fe1dc2 |
| 5 | 6d4b2f3c-9767-40ac-b064-c299b1032495 | 09/12/23 | 1218204 | 350e7879-7d0f-48e8-b8e6-1cf7b3d026a6 |
| 6 | e36a3b0a-7ef7-4f88-a98e-1cf7b3d026a6 | 23/11/23 | 2119983 | 350e7879-7d0f-48e8-b8e6-1cf7b3d026a6 |
| 7 | 62d4e7b2-7992-41cc-a8d4-90f5d29092d3 | 08/12/23 | 239843 | 903019f0-1261-4d44-a1d9-b91b51be12c6 |
| … | … | … | … | … |
| 4.100 | f15b7daf-63b6-4def-8db8-5e661134947f | 08/10/23 | 1218204 | af740747-1824-4cd2-a555-c62ebf9b08e0 |

*Deployment*

After completing the data according to RFM, then the program is displayed in the form of a web application. The web application developed uses a framework called streamlit and can be accessed at the address: https://segmentasi.streamlit.app/ .

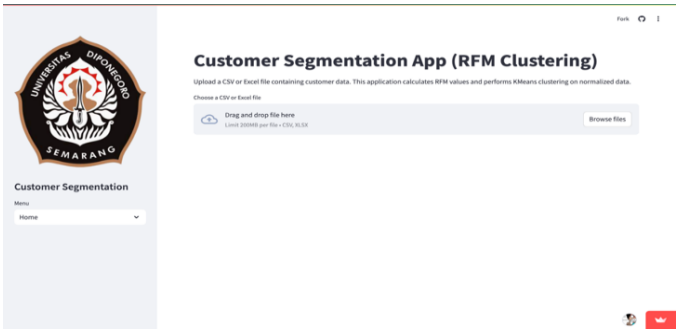


**Figure 1.** Web Application

Along with the successfully uploaded data, the main page will display the RFM Calculation analysis process in accordance with the data at the time of uploading at the beginning. The following is the result of normalizing the RFM value, into decimal numbers from each RFM dataset. It contains Customer ID, recency value, frequency value, and monetary value.

Customer segmentation data that has been formed, the result is a grouping of data that has the same data characteristics. From the data that has become the result of segmentation in the web application, the average value of each column or feature will be calculated automatically. Then from each existing cluster will be sorted based on the quality of the cluster. The quality of the web application will use the total of each column, but due to the different nature of each column, it is necessary to adjust the data by weighting. The following is an image of the weighting results and average RFM for each segment.

**Table 3.** RFM weighting and averaging

| Cluster | Customer Segment | Member Count | Recency (avg) | Frequency (avg) | Monetary (avg) | RFM Score |
|---|---|---|---|---|---|---|
| 0 | Frequent Buyers | 201 | 0.5784 | 0.1493 | 0.1734 | 0.3018 |
| 1 | High Spenders | 326 | 0.1413 | 0.504 | 0.4696 | 0.6053 |
| 2 | VIP Customers | 473 | 0.1575 | 0.1575 | 0.1947 | 0.4132 |

This weighting is done to prioritize what columns have more priority in value. The monetary and frequency columns will be good if the value is getting bigger, but for the recency column, it will be good if the value is getting closer to 0. Therefore, the weight value for recency is negative. Based on the total weighting value, the order of customer segmentation from the most to the least is as follows:

**Table 4.** Result clusters and customer segments

| Cluster | Customer Segment | Member | RFM Score |
|---|---|---|---|
| 2 | High Spenders | 326 | 0.6053 |
| 0 | VIP Customers | 473 | 0.4132 |
| 1 | Frequent Buyers | 201 | 0.3018 |

Based on Table 4, 326 customers in cluster 2 are included in the high spenders customer class with an RFM score of 0.6053. After the order is successfully obtained, then to facilitate the analysis process at the next stage, each of these clusters will be given a customer class and treatment according to purchasing behavior.

**Table 5.** Segmentation by cluster

| Cluster | Customer Segment | Member Count | RFM Score | Behavior | Treatment |
|---|---|---|---|---|---|
| 2 | High Spenders | 326 | 0.6053 | The most valuable and loyal customers with high transactions | Provide personalized gifts, loyalty programs, and premium customer services |
| 0 | VIP Customers | 473 | 0.4132 | Customers who frequently transact, although their purchase value is not always high | Offer loyalty rewards and special promotions to increase purchase frequency |
| 1 | Frequent Buyers | 201 | 0.3018 | Customers with high transaction value but infrequent purchases | Upsell premium products or offer exclusive experiences |

It can be concluded that customers with the highest monetary and frequency are high spenders.

*Elbow Method*

To find the optimal value using the Elbow Method, each data will be searched for the SSE (Sum Square Error) value which has been automatically calculated in the Customer Segmentation App (RFM Clustering) application to make comparisons on each existing cluster. The results of the SSE of each data will be visualized to determine which cluster has the optimal value.

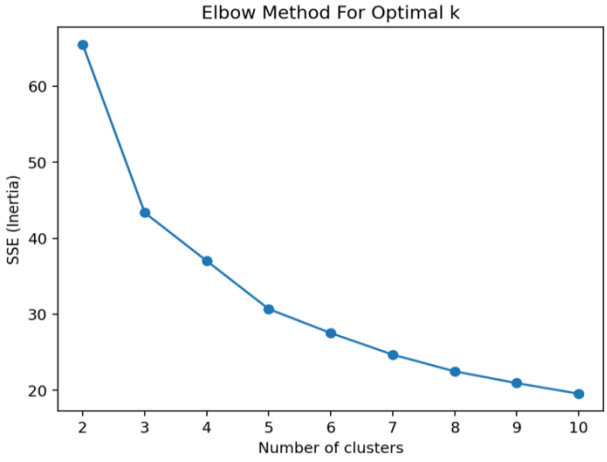


**Figure 2.** Curved Elbow Result

From the results of the application of the Elbow Method, the cluster with the optimal value is in cluster 3 which is the point with the perfect SSE value slope. So in the application of the Elbow Method it can be concluded that the optimal value is as many as 3 clusters.

*Silhouette Coefficient*

In finding the best cluster results in the application of the K-means algorithm, an iterative analysis stage is carried out to determine the best cluster results. The customer clustering evaluation method used in this web application is the Silhouette Coefficient method.

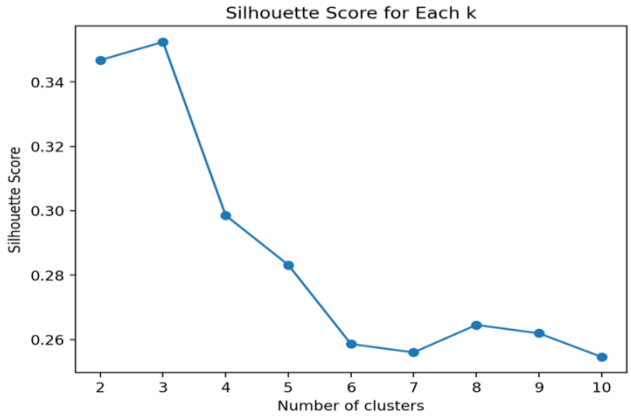


**Figure 3.** Silhouette Coefficient Result

As a result of the evaluation using this method, the highest value customer group is cluster 3 with a score of 0.3524.

*Comparison of RFM Average Values*

From the comparison of the three RFM Score Graph results, it can be seen that the colors used in each cluster illustrate a new status. Starting from the orange color which indicates that cluster group 2 has a high spender status, the blue color indicates that the customer has VIP Customer status and the gray color gives the status of the frequent buyers customer class.


**Figure 4.** RFM Graph Score

Based on the RFM score results in the table graph, cluster 2 with 473 customers has an average monetary value of 0.4696, frequency value of 0.504 and recency value of 0.413. So, the characteristics of cluster 2 customer class are high spenders, namely the most valuable and loyal customers with high transactions. Treatment suggestions that can be given are giving gifts to each customer, loyalty programs, and premium customers.

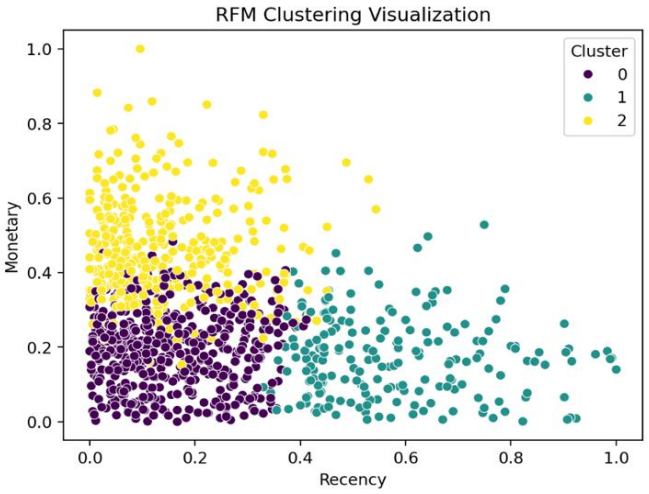


**Figure 5.** Scatter Plot RFM Clustering

The scatter plot shows the distribution points of each data resulting from the clustering process. Each cluster has different characteristics. The yellow dot indicates cluster 2, the purple dot is cluster 0 and the green dot is for cluster 1.

## Conclusion

This study explores customer segmentation using the RFM model and K-Means clustering: (1) The RFM model effectively describes customer characteristics based on Recency, Frequency, and Monetary values; (2) K-Means clustering is a viable method for segmentation, but manual computation is time-consuming. The Customer Segmentation App (RFM Clustering) automates this process, integrating key calculations; (3) Testing shows that three clusters are optimal, with Cluster 2 being the most profitable based on Frequency and Monetary values (Silhouette score: 0.3524); (4) This research contributes to data-driven customer segmentation, validated by the Elbow Method and Silhouette Coefficient; (5) Strategic recommendations include developing a web-based tool for real-time analysis, optimizing marketing efforts, and using segmentation data to assess campaign effectiveness.
This study helps businesses improve customer management and serves as a reference for future research.

**Acknowledgments**
Thank you to everyone who has helped and supported the implementation of this research. Without the assistance and guidance from various parties, this research would not have been successfully completed. I hope the results of this research can be beneficial and make a positive contribution in the future.

**Author Contributions**
Conceptualization, web application development, methodology, preliminary design, formal analysis, investigation, and visualization, E. S. Writing review and editing, validation, supervision, and resources, B. S. and D. M. K. N.

**Funding**
This research was funded by the researcher himself.

**Conflicts of Interest**
The authors declare no conflict of interest.

## References

Alam. A., Ahamad. M. K. (2024). K-Means Hybridization with Enhanced Firefly Algorithm for High-Dimension Automatic Clustering. *Journal of Advance Research in Aplied Sciences and Engineering Technology*. 33(3). 137-153. https://doi.org/10.37934/araset.33.3.137153

Ashari, F., Ilham, Nugroho, E., Baraku, R., Yanda, I., & Liwardana, R. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. *Journal of Applied Informatics and Computing*. 7, 89-97. https://doi.org/10.30871/jaic.v7i1.4947

Awalina, E. F L., & Rahayu. W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means *Clustering* pada Transaksi Online Retail. *Jurnal Mahasiswa Teknik Informatika (JATI)*. 13(2). https://doi.org/10.34010/jati.v13i2

Bagul. N., Surana. P.P., Berad. P., Khachane. C. (2021). Retail Customer Churn Analysis Using RFM Model and K-Means Clustering. *International Journal of Engineering Research & Technology*. 10(3). Retrieved from https://www.ijert.org/retail-customer-churn-analysis-using-rfm-model-and-k-means-clustering

Christy. A. J., Umamakeswari. A., Priyatharsini. L., and Neyaa. A. (2021). RFM ranking-an effective approach to customer segmentation. *Journal of King Saud University-Computer and Information Sci-ences*. 33 10, https://doi.org/10.1016/j.jksuci.2018.09.004

Firmansyah, H., Widyana, S. (2021). Pengaruh Customer Relationship Management (CRM) Terhadap Kepuasan Pelanggan Produk Sepatu Converse. *Jurnal Bisnis dan Pemasaran*. 11(1). Retrieved from https://ejurnal.ulbi.ac.id/index.php/promark/article/view/1321

Hayati, S., Suroso, A., Suliyanto, & Elfan Kaukab, M., (2020). Customer satisfaction as a mediation between micro banking image, customer relationship and customer loyalty. *Management Science Letters*, 10(11), 2561–2570. http://dx.doi.org/10.5267/j.msl.2020.3.039

Hidayanti. R., Zubair. A., Pratama. A. H., Indana. L. (2021). Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering. *Indonesia Journal on Computer and Information Technology. Techno.COM*, 20(2), http://dx.doi.org/10.33633/tc.v20i2.4556

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., Heming, J., & Abualigah, L. (2023). K-means *clustering* algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178-210. https://doi.org/10.1016/j.cosrev.2024.100621

Khan. I. K., Daud. H. B., Zinuddin. N. B., Sokkalingam. R., Farooq. M., Baig. M. E., Ayub. G., Zafar. M. (2024). Dermining the optimal number of clusters by Enhanced Gap Statistic in K-mean algorithm. *Egyptian Informatics Journal*. 27(1) https://doi.org/10.1016/j.eij.2024.100504

Kong. W., Wang. Y., Dai. H., Zhao. L., Wang. C. (2021). Analysis of Energy Consumption Structure Based on K-Means Clustering Algorithm. *E3S Web Conference*. 267, 01054. https://doi.org/10.1051/e3sconf/202126701054

Kristanti.B. T., Junaidi. A., Mandyartha. E. P. (2024). Implementasi K-Means *Clustering* Dalam Segmentasi pelanggan Berdasarkan Usia, Pendapatan, dan Model RFM (studi kasus: Lantikya Store Jomabang). *Jurnal Informatika dan Teknik Elektro (JITET)*. 12(3). http://dx.doi.org/10.23960/jitet.v12i3.4677

Maskanah. I., Primajaya. A., Rizal. A. (2020). Segmentasi Pelanggan Toko Purnama dengan Algoritma K-Means dan Model RFM untuk Perancangan Strategi Pemasaran. *Jurnal Inovtek Polbeng*. 5(2). https://doi.org/10.35314/isi.v5i2.1443

Nikmatun, I. A. (2019). Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*. 10(2). https://doi.org/10.24176/simet.v10i2.2882

Nisa. K., Heikal. J. (2022). Strategi Sa Pelanggan Manja Beauty Skincare Dengan Menggunakan Analisa Rfm Model. *Jurnal Mahasiswa Teknik Informatika (JATI)*. 6(1). https://doi.org/10.36040/jati.v6i1.4558

Rocha. L. M., Zela. M. A .C ., Torres. N. I. V., Medina. G. S. (2021). Analogy of the Application of Clustering and K-Means Tehniques for the Approxinimation of Values of Human Development Indicators. *International Journal of Advanced Computer Science and Applications*. 12(9). http://dx.doi.org/10.14569/IJACSA.2021.0120959

Sarkar. M., Puja. A. R., Chowdhury. F. R. (2024). Optimizing Marketing Strategies with RFM Method and K-Means Clustering Based AI Customer Segmentation Analysis. *Journal of Business and Management Studies* 6(2):54-60. http://dx.doi.org/10.32996/jbms.2024.6.2.5

Shirole. R., Salokhe. L., Jadhav. S. (2021). Customer Segmentation using RFM Model and K-Means Clustering. *International Journal of Scientific Research in Science and Technology*. 8(3). https://doi.org/10.32628/IJSRST2183118

Vysala. A., Gomes. D. J. (2020). Evaluating and Validating Cluster Results. *Computer Science & Information Technology (CS & IT)*. 37-47. https://doi.org/10.5121/csit.2020.100904